VISIT…

Chapter 3

Surface Topography

Symbols

$\bar{b}$	intercept on z axis, m
$\bar{d}$	difference in two methods of evaluating z, m
f	friction force, N
h	film thickness, m
$h_{\min}$	minimum film thickness, m
ℓ^*	bearing length, m
M_n	nth moment of distribution curve $\psi(z)$ about mean
$\bar{m}$	slope
N	number of measurements
R_a	centerline average or arithmetic average, m
R_q	root-mean-square (rms) surface roughness, m
R_t	maximum peak-to-valley height, m
w_z	normal load component, N
x	Cartesian coordinate in the direction of sliding, m
z	Cartesian coordinate in the direction of film, m
$\bar{\alpha}$	skewness
$\bar{\beta}$	kurtosis
Λ	dimensionless film parameter
μ	coefficient of sliding friction
$\tilde{\rho}_k$	autocorrelation
$\bar{\sigma}$	standard deviation, m
$\tilde{\psi}$	probability density function, m^{-1}

Subscripts

a	solid a
b	solid b
x, y, z	coordinate

3.1 Introduction

Increasing production speeds and new manufacturing methods, such as plasma cutting, spark erosion, chemical-mechanical polishing, and laser cutting, change the characteristics of machined surfaces. Requirements with regard to surface accuracy and surface refinement have also greatly increased. The importance of a fine-scale surface description is well demonstrated in tribology. The breakdown of lubrication layers of oil in engine cylinders or in bearings can be caused

Copyright © 2004 Marcel Dekker, Inc.

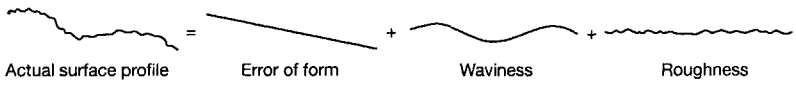


Figure 3.1: Geometric characteristics of solid surfaces [*From Halling (1976).*]

by improper microlevel surface shape.

Although bearing design theory relies heavily on fluid mechanics and kinematics, it is still ultimately a problem of two surfaces that are either in contact or separated by a thin fluid film. In either case the surface texture can be important in ensuring proper lubrication.

The first step in gaining insight into the lubrication of solid surfaces is to examine the surface profile, or topography. Smooth surfaces are not flat on an atomic scale. The roughness of manufactured surfaces used in lubrication is between 0.01 and 10 μm, whereas typical atomic diameters are between 0.0001 and 0.001 μm. Even a highly polished surface, when examined microscopically or with a profilometer, has an irregular nature. The surface consists of high and low spots. The high spots, or protuberances, are called "asperities."

Surface texture measurement and evaluation can be quite complex; only a few surface parameters and measurement issues are discussed here. Standardized nomenclature is contained in industry standards ASME B46.1 (ASME, 1995a) and ASME Y14.36M (ASME, 1995b).

A word of caution should be given that terms such as "rough", "fine", "smooth" and "super-smooth" should be avoided in describing the surface topography since the meaning is relative to the application being used. For example, a surface with R_q = 40 nm is considered "very rough" by people working in the field of optics, whereas a R_q=40 nm is considered "very smooth" for a machined surface.

3.2 Geometric Characteristics of Surfaces

The geometric characteristics, or texture, of surfaces as shown in Fig. 3.1 may conveniently be divided into three main categories:

1. *Error of form.* The surface deviates from a well-defined pattern because of errors inherent in the manufacturing process.

2. *Waviness.* Relatively long waves in a surface profile are often associated with unwanted vibrations that always occur in machine tool systems.

3. *Roughness.* Irregularities, excluding waviness and error of form, are inherent in the cutting and polishing process during production.

In the study of lubricated surfaces, roughness is the geometric variation that is generally of interest. Although often no sharp distinction can be drawn between

Copyright © 2004 Marcel Dekker, Inc.

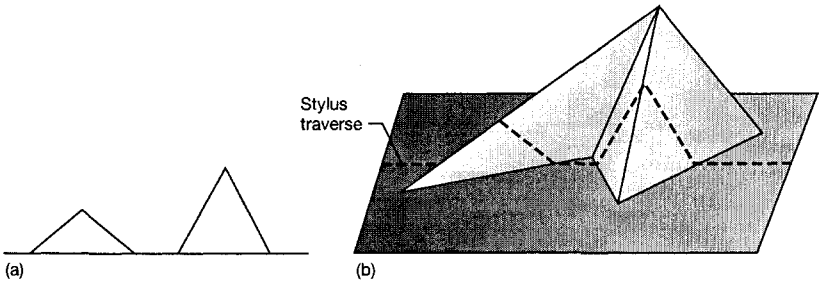


Figure 3.2: Difficulty in interpreting profilometer traces. (a) Surface profile; (b) surface asperity.

these categories, roughness simply concerns the horizontal spacing (wavelength) of the surface features. From a practical point of view, in characterizing surfaces used in tribology both the vertical direction (or amplitude parameter) and the horizontal direction (or wavelength) are important.

3.3 Contacting Measurement Methods

Two general classes of hardware are commonly used for measuring surface finish: contacting methods using stylus techniques and noncontacting methods. Contacting measurements are discussed in this section. Contacting measurements are based on transforming the vertical motion of a stylus tip as it traverses a surface into an electrical analog voltage. This voltage is then processed either by using analog circuitry or by converting it to digital information for processing. The method was introduced by Abbot and Firestone in 1933

3.3.1 Stylus Profilometry

Stylus profilometry is by far the most common method of measuring surface roughness because of the flexibility and comparatively low cost of profilometers. The stylus is normally made of diamond and has a tip radius of 2 μm and a static load of less than 0.0007 N (0.00256 oz). The tip radius is relatively large in comparison with the typical roughness. Therefore, it is often difficult to obtain a true picture of a surface from a stylus measurement. Features that appear to be asperity peaks on a single profile may in fact be local ridges on the flank of a true summit, as shown in Fig. 3.2. Moreover, many real surfaces of practical interest to tribologists have highly anisotropic surface textures. Thus, profiles taken in different traverse directions look quite different.

However, the main limitation of stylus measurements is the finite size of the stylus tip, which distorts the surface profile, as shown in Fig. 3.3, broadening peaks and narrowing valleys. The magnification in the vertical direction is

Copyright © 2004 Marcel Dekker, Inc.

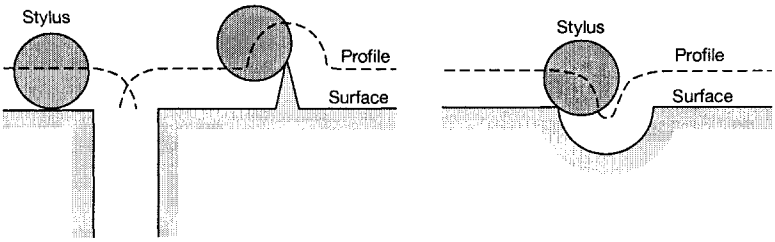


Figure 3.3: Error due to stylus radius.

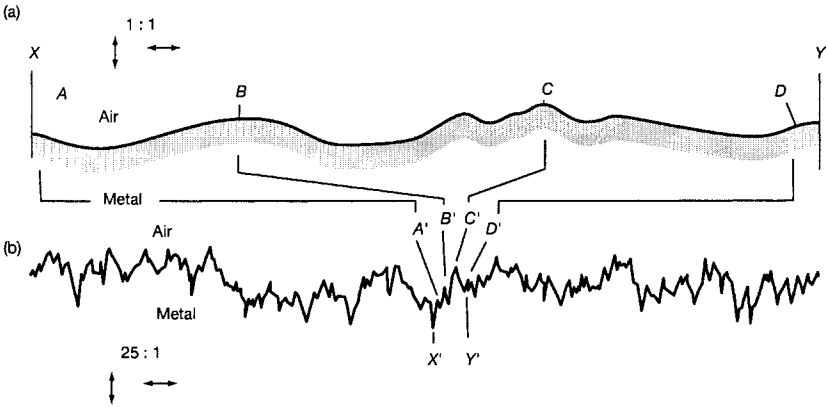


Figure 3.4: True (a) and compressed (b) profile. [*From Thomas (1982).*]

generally 100 to 100,000 and that in the horizontal direction is 10 to 5000. A typical ratio of vertical to horizontal magnification is 50:1. It is therefore important to take into account this difference in magnification. This is often not done, and thus a false impression of the nature of surfaces has become prevalent. Figure 3.4 attempts to illustrate this. In Fig. 3.4 the shape of a real surface is shown in the (a) portion of the figure and is compared with profilometer readings of the same surface in the (b) portion of the figure at a magnification ratio (vertical to horizontal) of 25:1.

Profilometers can measure single traces of surfaces, or else they can index across a surface to obtain a three dimensional contour.

3.3.2 Atomic Force Microscopy

Very smooth surfaces are difficult to characterize by standard stylus profilometry because they lack the vertical resolution and because of stylus size issues.

Copyright © 2004 Marcel Dekker, Inc.

An instrument available for the characterization of three dimensional, exceptionally smooth surfaces is the atomic force microscope (AFM). There are a number of AFM designs, but each involves moving the cantilever or specimen in raster fashion and recording the height distribution of a surface. As shown in Fig. 3.5, a laser is focused onto the end of the cantilever and then reflected onto a series of four photodetectors. The position of the cantilever end can then be accurately measured by comparing the voltage registered by each photodetector.

There are two common modes for surface measurement in an AFM: contact mode and tapping mode. In contact mode, the sample and cantilever are displaced relative to one another and the associated cantilever displacement yields a surface profile. In tapping mode, the cantilever is vibrated to near its natural frequency, and it is placed close enough to the surface so that it taps the sample. The surface height profile is then recorded as the maximum vertical deflection of the cantilever. Typical cantilevers used in AFM surface measurements are shown in Fig. 3.5b. Further information regarding the utility of AFM operation can be found in Bhushan (1998, 2002).

3.4 Noncontacting Measurement Devices

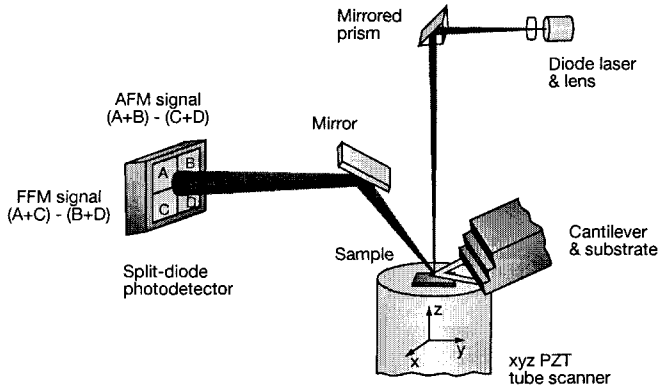
The noncontacting methods for determining the characteristics of surfaces use a number of different measurement principles and devices:

1. *Laser-based devices.* With these devices, a sample is placed under the sensor and is accurately positioned and rastered while the sensor transmits the height data to the measurement control unit. The sensor uses light focused from a laser diode onto the specimen surface. The reflected light is directed to a focus detector, which measures deviations from the ideal focus to within a few nanometers. The deviation in focus generates an error, which is used to refocus. The position of the objective represents an absolute measurement of the height.

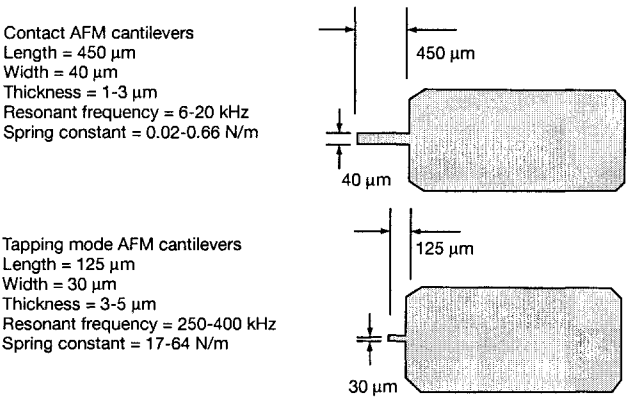
2. *Pneumatic devices.* This method is based on the measurement of air leakage. A flat and finely finished measurement head has an orifice of air. Measuring the airflow when the head is placed on the surface indicates changes in surface roughness. Pneumatic devices provide a simple, inexpensive, portable, quick, and robust way of assessing surface roughness, well-suited for use in determining the quality of surfaces on the shop floor.

3. *Optical devices.* The intensity of reflected laser light defines the surface texture, producing an average over a surface area. The result is often called gloss, and attempts to correlate gloss to surface roughness have been only partially successful. This method is not well established in terms of the number of people who use it routinely.

Copyright © 2004 Marcel Dekker, Inc.



(a)


Contact AFM cantilevers
Length = 450 μm
Width = 40 μm
Thickness = 1-3 μm
Resonant frequency = 6-20 kHz
Spring constant = 0.02-0.66 N/m
450 μm
40 μm
Tapping mode AFM cantilevers
Length = 125 μm
Width = 30 μm
Thickness = 3-5 μm
Resonant frequency = 250-400 kHz
Spring constant = 17-64 N/m
125 μm
30 μm
Material: Etched single-crystal n-type silicon;
resistivity = 0.01-0.02 ohm/cm
Tip shape: 10 nm radius of curvature, 35° interior angle

(b)

Figure 3.5: Schematic illustration of an atomic force microscope. (a) Principle of operation for an AFM. (b) Typical cantilevers for use in an atomic force microscope. [*Source: Digital Instruments Corp.*]

Copyright © 2004 Marcel Dekker, Inc.

4. *Electron microscope devices.* This method offers better resolution and depth of field than optical methods, chiefly because of the extremely short wavelength of electron microscope beams in comparison with that of light. Sherrington and Smith (1988) point out the two types of electron microscopy:

   (a) *Transmission electron microscopy* (TEM). Electrons are incident on a thin specimen (less than 10 nm thick) that deflects and scatters the electrons as they pass through it. A lens system magnifies and focuses scattered electrons to form an image on a screen or photographic film. To examine the surface of a metal component by TEM, it is usually necessary to fabricate a replica of the specimen surface. TEM is typically able to resolve features down to a separation of around 0.3 nm and has been used to study the changes in surface structure during wear.

   (b) *Reflection electron microscopy* (REM). Electrons are scattered from the surface of the specimen and strike a collector, which generates an electrical signal. This signal is subsequently processed and used to form an image representing the specimen surface on a monitor screen. Scattered electrons are produced by a beam of finely focused electrons that scans the specimen in a raster pattern, a process called "scanning electron microscopy" (SEM). Scanning electron microscopes can be adjusted to have a maximum resolution of about 10 nm, a little less than that available in TEM. However, this disadvantage is compensated for by the fact that specimen preparation is considerably easier.

Besides these noncontacting devices, scanning tunneling microscopy is also available but beyond the scope of this book. Sherrington and Smith (1988) give an excellent description of this device as well as describe in general the modern measurement techniques used in determining the characteristics of tribological surfaces.

Several methods of measuring surface topography have been described in this section as well as in the preceding section. Of these methods the stylus measurement device is the most widely used. Table 3.1, obtained from Sherrington and Smith (1988) and Stout et al. (1993), summarizes the specifications of a number of the devices described in this section as well as in the preceding section. From this table the resolution, the depth of field, and the measurable area are described for four different devices.

3.5 Reference Lines

In computing the parameters that define the surface texture all height measurements are made from some defined reference line. Several methods have been used. They are summarized below and shown graphically in Fig. 3.6.

Copyright © 2004 Marcel Dekker, Inc.

Table 3.1: Summary of typical specifications of devices used for surface topography measurement. From Sherrington and Smith (1988).

Device	Resolution				Vertical measurement range or depth of field		Measurable area
	Lateral		Vertical		At lowest resolution	At highest resolution	
	Lowest	Highest	Lowest	Highest			
Stylus instrument	a	a	0.5 μm	0.25 nm	500 μm	0.25 μm	Depends on traverse length; typically a few millimeters
Optical light microscope	2.5 μm	0.1 μm	b	b	42 μm	0.04 μm	Depends on magnification
Transmission electron microscope	2.5 nm	0.5 nm	c	c	400 nm	80 nm	Depends on magnification
Scanning electron microscope	5 μm	10 nm	d	d	1 mm	2 μm	Depends on magnification

[a] Not easily defined.
[b] Not applicable.
[c] Approximately the same as the lateral resolution.
[d] Not available.

Copyright © 2004 Marcel Dekker, Inc.

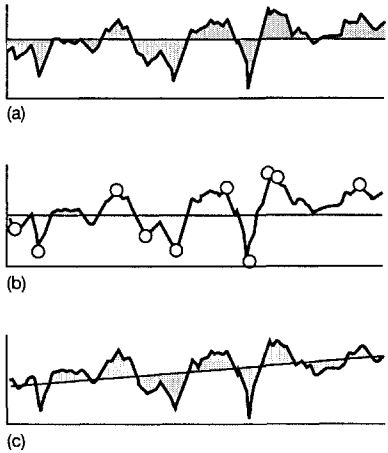


Figure 3.6: Comparison of three types of reference line: (a) M system; (b) ten-point average; (c) least squares.

3.5.1 Mean, or M System

The mean, or M system, method is based on selecting the mean line as the centroid of the profile. Thus, the areas above and below this line are equal. If for discrete profiles the area of each profile is a rectangle, this method turns out to be simply finding the average of the measured heights. This method gives a "horizontal" reference that does not compensate for errors of form or tilt.

3.5.2 Ten-Point Average

The ten-point average method is based on finding the five highest peaks and the five lowest valleys. The average of these 10 points gives the reference line. For deeply pitted surfaces this method can lead to a reference line that is below the major surface features.

3.5.3 Least Squares

The least-squares method is based on postulating a sloping reference line instead of a horizontal line as is the case for the M system. Therefore, the major advantage of this approach is that it can be used to compensate for the linear error of form, or tilt. The following will attempt to describe mathematically the least-squares reference lines. From the equation of a line,

$$z = \bar{m}x + \bar{b}$$

where

Copyright © 2004 Marcel Dekker, Inc.

$\bar{m}$ = slope
$\bar{b}$ = intercept on z axis, m

Given a set of points $P_1(x_1, z_1), P_2(x_2, z_2), \ldots, P_n(x_n, z_n)$ corresponding to each value of x, consider two values of z: (1) z measured and (2) z obtained from $\bar{m}x + \bar{b}$. Call the difference $\bar{d}$ such that

$$\bar{d}_1 = \left[z_1 - \left(\bar{m}x_1 + \bar{b}\right)\right], \ldots, \bar{d}_n = \left[z_n - \left(\bar{m}x_n + \bar{b}\right)\right]$$

The set of all deviations gives a picture of how well the observed data fit a line. If $\sum_{i=1}^{n} \bar{d}_i^2 = 0$, the fit is perfect. This never occurs in real situations, and that is where the method of least squares comes in.

Not only is it important to establish a reference line or mean line in the z direction, it is also important to establish a sampling length or distance in the x direction that adequately differentiates between roughness and waviness.

$$\bar{f}\left(\bar{m}, \bar{b}\right) = \sum_{i=1}^{n} \bar{d}_i^2$$

or

$$\bar{f}\left(\bar{m}, \bar{b}\right) = \left(z_1 - \bar{m}x_1 - \bar{b}\right)^2 + \left(z_2 - \bar{m}x_2 - \bar{b}\right) + \cdots + \left(z_n - \bar{m}x_n - \bar{b}\right)^2$$

Finding the values of $\bar{m}$ and $\bar{b}$ that give the smallest $\bar{f}\left(\bar{m}, \bar{b}\right)$ requires finding values of $\bar{m}$ and $\bar{b}$ that satisfy

$$\frac{\partial \bar{f}}{\partial \bar{m}} = \frac{\partial \bar{f}}{\partial \bar{b}} = 0$$

Solving these two equations for the two unknowns produces the reference line that gives the slope-intercept form.

Once the method of selecting a reference line is agreed upon, all surface height measurements are made relative to it. The centroid (or M system) and the least-squares reference line methods are the easiest to deal with analytically, because the sum of all the profile height deviations is always zero.

3.6 Computation of Surface Parameters

Measurements made with a stylus instrument are assumed. The reference line used is obtained by either the M system or the least squares method so that the average of z_i is zero. Samples taken at uniform length intervals are assumed. The Δ is defined as small. Resulting discretized height values are denoted by $z_i, i = 1, 2, \cdots, N$.

Three different surface parameters may be computed:

1. Centerline average (CLA) or arithmetic average (AA), denoted by R_a,

$$R_a = \frac{1}{N} \sum_{i=1}^{N} |z_i| \tag{3.1}$$

Copyright © 2004 Marcel Dekker, Inc.

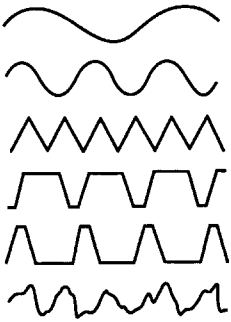


Figure 3.7: Geometric profiles having same values of arithmetic average. [*From Halling (1976).*]

From $\partial \bar{f}/\partial \bar{b} = 0$ it immediately follows that $\sum \bar{d}_i = 0$ so $z_i' = z_i - \bar{m}x_i - \bar{b}$ has zero mean.

2. Root mean square (rms), denoted by R_q,

$$R_q = \left(\frac{1}{N} \sum_{i=1}^{N} z_i^2 \right)^{1/2} \tag{3.2}$$

If a Gaussian height distribution is assumed, the R_q has the advantage of being the standard deviation of the profile.

3. Maximum peak-to-valley height, denoted by R_t,

$$R_t = \max(z) - \min(z) \tag{3.3}$$

In general,

$$R_a < R_q < R_t \tag{3.4}$$

Also for a simple sine distribution the ratio of R_q to R_a is

$$\frac{R_q}{R_a} = \frac{\pi}{2\sqrt{2}} = 1.11 \tag{3.5}$$

Table 3.2 gives typical values of the arithmetic mean R_a for various processes and components. Figure 3.7 shows six different surface profiles, all with the same R_a, or arithmetic average roughness. The R_a is thus an ambiguous parameter, since it does not indicate whether the R_a value is the mean of many small deviations from the mean value or of a few large ones. For this reason tribologists seek additional surface parameters that are rather more informative.

Thus far, the discussion of surface roughness has been on defining the profile in the z direction. To incorporate the length of the asperities, the bearing length is introduced. The profile bearing length is obtained by cutting the profile peaks by a line parallel to the mean line within the sampling length at a given section

Copyright © 2004 Marcel Dekker, Inc.

Table 3.2: Typical arithmetic averages for various processes and components.

	Arithmetic average R_a	
	μm	μin.
Processes		
Sand casting; hot rolling	12.5–25	500–1000
Sawing	3.2–25	128–1000
Planing and shaping	0.8–25	32–1000
Forging	3.2–12.5	128–500
Drilling	1.6–6.3	64–250
Milling	0.8–6.3	32–250
Boring; turning	0.4–6.3	16–250
Broaching; reaming; cold rolling; drawing	0.8–3.2	32–128
Die casting	0.8–1.6	32–64
Grinding, coarse	0.4–1.6	16–64
Grinding, fine	0.1–0.4	4–16
Honing	0.1–0.8	4–32
Polishing	0.05–0.4	2–16
Lapping	0.025–0.4	1–16
Components		
Gears	0.25–10	10–400
Plain bearings - journal (runner)	0.12–0.5	5–20
Plain bearings - bearing (pad)	0.25–1.2	10–50
Rolling bearings - rolling elements	0.025–0.12	1–5
Rolling bearings - tracks	0.1–0.3	4–12

level. Figure 3.8 is helpful in describing the bearing length ℓ^*. Mathematically, it is defined as

$$\ell^* = \ell_1^* + \ell_2^* + \cdots + \ell_n^* \tag{3.6}$$

The profile bearing length ratio t_p is defined as

$$t_p = \frac{\ell^*}{\ell} \tag{3.7}$$

where ℓ = sampling length, m.

The Abbot curve (also known as the "bearing length curve") is shown in Fig. 3.9. The vertical axis is the bearing length ratio multiplied by 100. These curves describe how much of the profile protrudes a given distance above the surface. During running-in, a certain amount of the surface will be removed by plastic deformation. These curves also calculate the peakedness and the amount of material between given heights. In the top profile the asperities have steeper peaks than in the bottom profile.

The texture of a surface can be described in terms of the distribution function of its profile heights. In statistical terms the cumulative distribution of the all-ordinate distribution curve can be written as

$$\bar{F}(z) = \int_{-\infty}^{z} \tilde{\psi} dz$$

Copyright © 2004 Marcel Dekker, Inc.

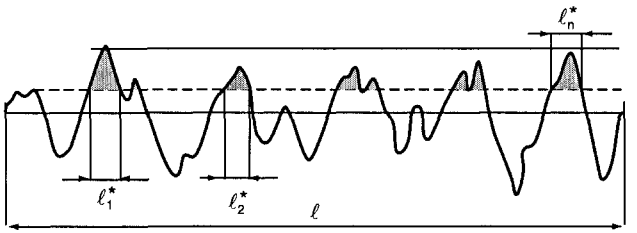


Figure 3.8: Surface profile showing bearing length. [*From Persson (1992).*]

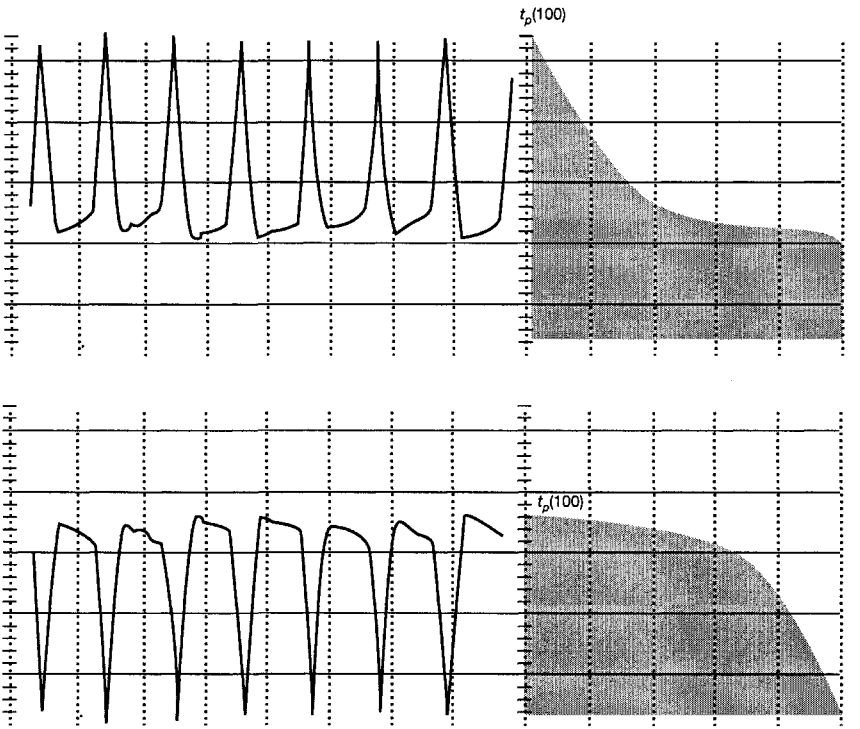


Figure 3.9: Abbot curves for two different profiles. [*From Persson (1992).*]

Copyright © 2004 Marcel Dekker, Inc.

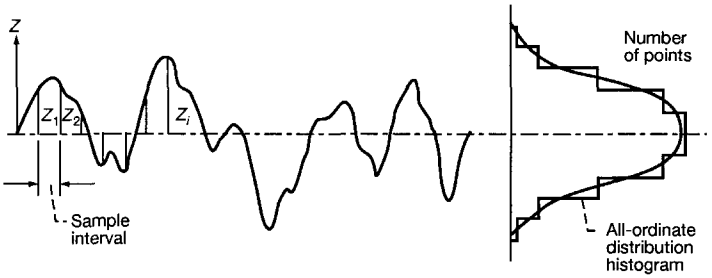


Figure 3.10: Method of deriving all-ordinate distribution. [*From Halling (1975).*]

where z refers to the profile height and $\tilde{\psi}$ is the probability density function of the distribution of these heights. The probability density function may be viewed as the fraction of heights in a given interval. Therefore, the practical derivation of such a distribution curve involves taking measurements of z_1, z_2, etc., at some discrete interval and summing the number of ordinates at any given height level. Figure 3.10 illustrates the method used in obtaining the all-ordinate distribution. The distribution curve in Fig. 3.10 is the smoothest curve that can be drawn through the histogram produced by such a sampling procedure. This smooth curve for many surfaces tends to exhibit a Gaussian distribution of surface texture heights.

If these ideas of height distribution are introduced, the empirical probability density, or histogram, can be used to evaluate R_a and R_q. If the fraction of heights in an interval $z_j - \Delta \leq z_j \leq z_j + \Delta$ is denoted by $\tilde{\psi}_j$ for $j = -L, \ldots, 0, \ldots L$, then approximately the same values as those given in Eqs. (3.1) and (3.2) can be expressed as

$$R_a = \int_{-L}^{L} |z| \tilde{\psi} dz \tag{3.8}$$

$$R_q = \left(\int_{-L}^{L} z^2 \tilde{\psi} dz \right)^{1/2} \tag{3.9}$$

From the Gauss-Laplace law a general expression for the probability density function while assuming a Gaussian distribution is

$$\tilde{\psi} = \frac{1}{\bar{\sigma} (2\pi)^{1/2}} \exp \frac{-(z - z^*)^2}{2\bar{\sigma}^2} \tag{3.10}$$

where $\bar{\sigma}$ is the standard deviation and z^* is the distance of the mean from the value chosen as the origin. The values of the ordinate $\tilde{\psi}$ of the Gaussian

Copyright © 2004 Marcel Dekker, Inc.

distribution curve for $z^* = 0$ are found in most books on statistics. The form of the Gaussian distribution necessitates a spread of $-\infty$ to ∞, which cannot happen with practical surfaces. In practice the distribution curve is truncated to $\pm 3\bar{\sigma}$. Because approximately 99.9% of all events occur within this region, the truncation leads to negligible error while providing useful simplification.

The nth moment of the distribution curve $\tilde{\psi}dz$ about the mean is defined as

$$M_n = \int_{-\infty}^{\infty} z^n \tilde{\psi} dz \tag{3.11}$$

It can be observed that twice the first moment of half $\tilde{\psi}dz$ is equivalent to the centerline average R_a defined in Eq. (3.8) or

$$R_a = 2\int_0^{\infty} z\tilde{\psi}dz = \text{twice the first moment of half } \tilde{\psi}dz$$

The first moment of the whole $\tilde{\psi}dz$ about the mean reference line is zero. Likewise, comparing the second moment of $\tilde{\psi}dz$ with the root mean square R_q expressed in Eq. (3.9) gives

$$R_q = \bar{\sigma} = \left(\int_{-\infty}^{\infty} z^2\tilde{\psi}dz\right)^{1/2} = \left(\text{second moment of } \tilde{\psi}dz\right)^{1/2}$$

The third moment of $\tilde{\psi}dz$ relates to the skewness of a curve, or the departure of a curve from symmetry. The mathematical expression for the normalized skewness is

$$\bar{\alpha} = \frac{1}{R_q^3}\int_{-\infty}^{\infty} z^3\tilde{\psi}dz \tag{3.12}$$

If the peaks and valleys deviate from the reference line by approximately the same amount, the skewness will be zero. If, on the other hand, the surfaces are deeply pitted, the skewness will be some negative value.

The fourth moment of $\tilde{\psi}dz$ relates to the peakedness, or kurtosis, of the curve and is expressed as

$$\bar{\beta} = \frac{1}{R_q^4}\int_{-\infty}^{\infty} z^4\tilde{\psi}dz \tag{3.13}$$

The kurtosis always has a positive value and measures the sharpness of a symmetric distribution. For a Gaussian distribution the curve has a kurtosis of 3. When most of the profile heights are close to the reference line, $\bar{\beta}$ is quite large; a relatively flat height distribution will have a $\bar{\beta}$ near zero. Curves with values of $\bar{\beta}$ less than 3 are called "platykurtic," and those with $\bar{\beta}$ greater than 3 are called "leptokurtic." Figure 3.11 illustrates the various types of kurtosis curves.

For a Gaussian distribution where the curve is represented by $\tilde{\psi}$, as expressed in Eq. (3.10), the general expression for the nth moment is given by

$$M_n = \frac{1}{\bar{\sigma}(2\pi)^{1/2}}\int_{-\infty}^{\infty} z^n \exp\frac{-z^2}{2\sigma^2}dz \tag{3.14}$$

Copyright © 2004 Marcel Dekker, Inc.

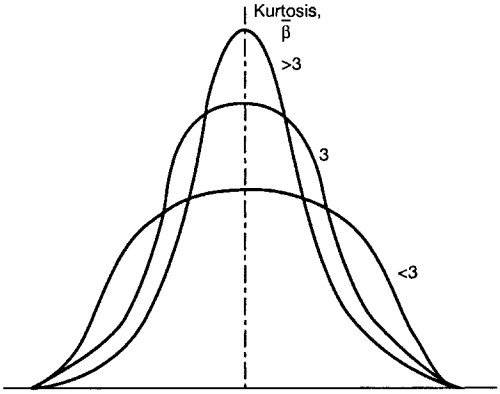


Figure 3.11: Illustration of three different kurtosis values. [*From Halling (1975).*]

where the standard deviation $\bar{\sigma} = R_q$. From Eq. (3.14) it is observed that, if n is odd, M_n vanishes and the curve must be symmetrical. If n is even, then

$$M_n = \frac{n!}{2^{n/2}\,(n/2)!}\bar{\sigma}^n \tag{3.15}$$

Note that the second moment becomes $\bar{\sigma}^2$, the variance.

$$\therefore M_2 = \bar{\sigma}^2 = R_q^2 = \text{variance} \tag{3.16}$$

Some additional parameters used to define surfaces utilized in fluid film lubrication as well as the range of values normally found are listed below:

Density of asperities, 10^2 to 10^6 peaks/mm^2
Asperity spacing, 1 to 75 μm
Asperity slopes, 0 to 25°, but mainly 5 to 10°
Radii of peaks, mostly 10 to 30 μm

These additional surface definitions better define the surface texture used in fluid film lubrication.

Fractals have not been used by researchers to characterize fluid film lubrication surfaces thus far, but they might be useful in the future. Fractals are geometric structures that have a similar appearance regardless of the scale on which they are being observed. If a surface profile exhibits self-simularity independent of the scales, the concept of fractal dimension can be used to characterize the profile.

Copyright © 2004 Marcel Dekker, Inc.

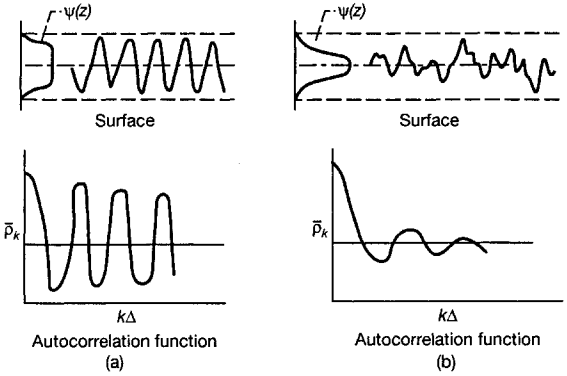


Figure 3.12: Two different surfaces and resulting autocorrelation functions. (a) Periodicity profile; (b) decay profile. [*From Halling (1975).*]

3.7 Autocorrelation Parameter

The previously discussed parameters (R_a, R_q, and R_t) depend only on the profile heights, not on the spacing between heights. As a result they do not truly reflect the shape of the profile, as illustrated in Fig. 3.10. The autocorrelation does incorporate the spacing between heights and is obtained by multiplying each profile height by the height of the point at some fixed horizontal distance farther along the profile. After averaging the product over a representative profile length and normalizing by the variance R_q^2, the expression for the autocorrelation is

$$\tilde{\rho}_k = \frac{1}{R_q^2\,(N-k)} \sum_{i=1}^{N-k} z_i z_{i+k} \tag{3.17}$$

The autocorrelation depends on k and is a measure of the similarity of heights separated by the distance $k\Delta$ (where Δ is the sample interval, assumed to be constant). In machining processes such as turning or milling, where there are pronounced feed marks, the autocorrelation will have a maximum where $k\Delta$ is some integer multiple of the linear feed. This distance is often called the "characteristic wavelength."

Typical plots of the autocorrelation function for two different profiles are shown in Fig. 3.12. The variance is defined as $\bar{\sigma}^2$. The shape of this function is most useful in revealing some of the characteristics of the profile. The general decay of the function is the random component of the surface profile and indicates a decrease in correlation as $k\Delta$ increases, as shown in Fig. 3.12b. The oscillatory component of the function indicates any inherent periodicity of the profile, as shown in Fig. 3.12a.

Copyright © 2004 Marcel Dekker, Inc.

3.8 Distribution of Slope and Curvature

A basic geometric description of a profile can be given, considering the profile as a random function, with first and second derivatives. Each of the derivatives also has a statistical distribution associated with it. The most straightforward way of estimating this information is to use differences in the profile to represent the derivatives, or

$$\left(\frac{dz}{dx}\right)_i = \frac{z_{i+1} - z_i}{\Delta} \tag{3.18}$$

$$\left(\frac{d^2z}{dx^2}\right)_i = \frac{z_{i+1} - 2z_i + z_{i-1}}{\Delta^2} \tag{3.19}$$

Equation (3.18) defines the slope of the profile and Eq. (3.19) defines its curvature. However, because the profile is a random function with "noise," differencing is not the best way, since it makes this noise greater. Some sort of smoothing of the data should be done to minimize this effect. One way of doing this is to use Eqs. (3.18) and (3.19) with samples taken at intervals of Δ, 2Δ, $3\Delta, \ldots$ and then to make weighted averages of these expressions. For example, by choosing weights of 1.5, –0.6, and 0.1 for derivatives formed at Δ, 2Δ, and 3Δ, respectively, the following expressions are obtained:

$$\left(\frac{dz}{dx}\right)_i = \frac{1}{60\Delta}\left[45\left(z_{i+1} - z_{i-1}\right) - 9\left(z_{i+2} - z_{i-2}\right) + \left(z_{i+3} - z_{i-3}\right)\right] \tag{3.20}$$

$$\left(\frac{d^2z}{dx^2}\right)_i = \frac{1}{180\Delta^2} \times \left[-490z_i + 270\left(z_{i+1} - z_{i-1}\right) - 27\left(z_{i+2} - z_{i-2}\right) + 2\left(z_{i+3} - z_{i-3}\right)\right] \tag{3.21}$$

A slope and curvature profile can be generated in this way from the height profile, and by using the same relations as for the heights, a mean or average slope and curvature can be found as well as the respective standard deviation of the distributions. Likewise, the skewness and the kurtosis can be computed to characterize the shape of these distributions.

The previously defined properties are a few of the parameters that can be found from analyzing a profile. These characteristics may be important in relating function to manufactured surfaces or in analyzing how a surface interacts with another surface. They illustrate some of the efforts at extracting more information from the surface measurement than simply the height sensitive parameters.

Finally, the relation between function and surface geometry has to be more fully understood. This will require closer interaction between the manufacturing engineer who makes the surface and the designer who specifies the surface.

Copyright © 2004 Marcel Dekker, Inc.

3.9 Film Parameters for Different Lubrication Regimes

If a machine element is adequately designed and fluid film lubricated, the lubricated surfaces are completely separated by a lubricant film. Endurance testing of ball bearings, for example, as reported by Tallian et al. (1967) has demonstrated that when the lubricant film is thick enough to separate the contacting bodies, the fatigue life of the bearing is greatly extended. Conversely, when the film is not thick enough to provide full separation between the asperities in the contact zone, the life of the bearing is adversely affected by the high shear resulting from direct metal-to-metal contact.

The four lubrication regimes were defined in Sec. 1.3, and calculation methods for determining the rms surface finish were introduced in Sec. 3.6. This section introduces a film parameter and describes the range of values for the four lubrication regimes. The relationship between the dimensionless film parameter Λ and the minimum film thickness $h_{\min}$ is

$$\Lambda = \frac{h_{\min}}{\left(R_{q,a}^2 + R_{q,b}^2\right)^{1/2}} \tag{3.22}$$

where

$R_{q,a}$ = rms surface finish of surface a
$R_{q,b}$ = rms surface finish of surface b

The film parameter is used to define the four important lubrication regimes. The range of Λ for these four regimes is

1. Hydrodynamic lubrication, $5 < \Lambda < 100$
2. Elastohydrodynamic lubrication, $3 < \Lambda < 10$
3. Partial lubrication, $1 < \Lambda < 5$
4. Boundary lubrication, $\Lambda < 1$

These values are rough estimates. The great differences in geometric conformity in going from hydrodynamically lubricated conjunctions to elastohydrodynamically lubricated conjunctions make it difficult for clear distinctions to be made.

Running-in is a process that affects the film parameter. This process allows wear to occur so that the mating surfaces can adjust to each other to provide for smooth running. This type of wear may be viewed as beneficial. The film parameter will increase with running-in, since the composite surface roughness will decrease. Running-in also has a significant effect on the shape of the asperities that is not captured by the composite surface roughness. With running-in, the peaks of the asperities in contact become flattened.

Copyright © 2004 Marcel Dekker, Inc.

Example 3.1

Given: Two equally rough surfaces are lubricated and loaded together ($R_{qa} = R_{qb}$). The film parameter is two.

Find: The value of Λ if one of the surfaces is polished so that $R_q \to 0$ (i.e., the surface becomes absolutely smooth).

Solution: From Eq. (3.22) the initial condition is

$$\Lambda_1 = \frac{h_{\min}}{\left(R_q^2 + R_q^2\right)^{1/2}} = \frac{h_{\min}}{\sqrt{2}R_q} = 2 \qquad (a)$$

When one of the surfaces becomes absolutely smooth,

$$\Lambda_2 = \frac{h_{\min}}{\left(R_q^2 + 0\right)^{1/2}} = \frac{h_{\min}}{R_q} \qquad (b)$$

$\therefore$ Λ increases from 2 to $2\sqrt{2}$.

Example 3.2

Given: Gears for an excavator are manufactured by sand casting. The surface roughness is measured to have a centerline average of 18 μm. This high surface roughness makes the gear wear rapidly. The film thickness for the grease-lubricated gears is calculated to be 1.6 μm.

Find: The method of manufacture for the gears to produce a film parameter of 10.

Solution: Using Eq. (3.22) while assuming that the roughnesses are equal on the two surfaces gives

$$\Lambda = \frac{h_{\min}}{R_q\sqrt{2}} \qquad \text{or} \qquad R_q = \frac{h_{\min}}{\Lambda\sqrt{2}} = \frac{1.6}{10\sqrt{2}} = 0.1131\mu m$$

Table 3.2 shows that for a surface roughness of 0.1 μm, grinding is the fastest and cheapest method of achieving these surface finishes. Smoother surfaces can be manufactured by honing, polishing, and lapping, but these processes are considerably more expensive and do not lend themselves well to the intricate geometry of gear teeth.

3.10 Transition Between Lubrication Regimes

As the severity of loading is decreased, there is no sharp transition from boundary to fluid film lubrication; rather an increasing proportion of the load is carried by pressures within the film that fills most of the space between the

Copyright © 2004 Marcel Dekker, Inc.

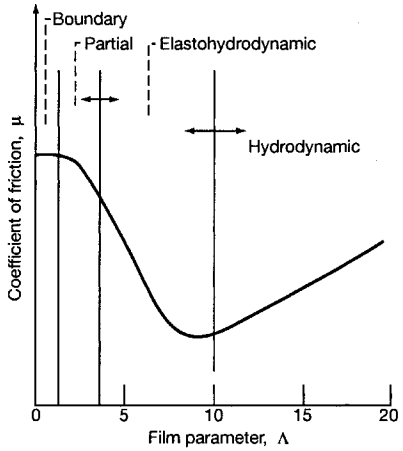


Figure 3.13: Variation of friction coefficient with film parameter [*From Hamrock and Dowson (1981).*]

opposing solids. Indeed, it is often difficult to eliminate fluid film lubrication effects so that true boundary lubrication can occur, and there is evidence to suggest that micro-fluid-film lubrication formed by surface irregularities is an important effect.

The variation of the friction coefficient μ with the film parameter Λ is shown in Fig. 3.13. This can be compared to the Stribeck curve of Fig. 1.16. The friction coefficient is defined as

$$\mu = \frac{f}{w_z} \tag{3.23}$$

where f is the tangential (friction) force and w_z is the normal applied load. In Fig. 3.13 the approximate locations of the various lubrication regimes discussed in Sec. 3.9 are shown. This figure shows that as the film parameter Λ increases, the friction coefficient initially decreases in the elastohydrodynamic regime and then increases in the hydrodynamic regime. In explaining this phenomenon let us assume the surface roughness is the same in both lubrication regimes.

In hydrodynamic lubrication of conformal surfaces as found in journal and thrust bearings, $w_z \propto 1/h^2$. In elastohydrodynamic lubrication of nonconformal surfaces the normal applied load has little effect on the film thickness. Therefore, w_z is essentially proportional to a constant. In both hydrodynamic and elastohydrodynamic lubrication the frictional force is due to lubricant shearing and can be expressed in both lubrication regimes as $f \propto 1/h$. Making use of this,

$$\mu_{HL} \propto \frac{1/h}{(1/h)^2} \propto h \tag{3.24}$$

$$\mu_{EHL} \propto \frac{1/h}{\text{constant}} \propto \frac{1}{h} \tag{3.25}$$

This then explains the reversing of the slope of the friction coefficient in Fig. 3.13.

Copyright © 2004 Marcel Dekker, Inc.

3.11 Closure

Since fluid film lubrication is concerned with the lubrication between solids separated by small film thicknesses, it is essential that the physical nature of the solid's surface topography be understood. To obtain this understanding, this chapter investigated surface measurement hardware. The contact method of stylus measurement is based on transforming the vertical motion of the stylus tip as it traverses a surface into an electrical analog voltage. The main limitation of this approach is the finite size of the stylus tip, which distorts the surface profile (broadening peaks and narrowing valleys). Atomic force microscopy allows more surface roughness measurement with higher resolution than profilometry. The following noncontacting measurement devices were presented.

1. Laser interferometers that move a lens in order to maintain focus on a sample surface. The laser can be focused at different locations in order to obtain a three-dimensional surface profile.

2. Pneumatic devices measure the airflow when a finely finished head is placed on the surface and indicate changes in surface roughness.

3. Optical devices use the intensity of reflected laser light to define the surface texture.

4. Electron microscopic devices produce images when electrons are incident on the specimen.

In computing the parameters that define the surface texture, all height measurements were made from some defined reference line. Two methods of defining the reference line are

1. Mean, or M system, based on selecting the mean line as the centroid of the profile. The area above this line equals the area below this line.

2. Least squares, based on postulating a slope-intercept form for the reference line. This can compensate for errors of form or tilt.

Methods of computing surface parameters were also discussed. The following two important parameters were presented:

1. Centerline average (CLA), or arithmetic average (AA), denoted by R_a.

2. Root mean square (rms) roughness, denoted by R_q. In general, $R_a < R_q$.

The film parameter Λ was defined as the ratio of the minimum film thickness to the composite surface roughness. The film parameter was used to define four important lubrication regimes: hydrodynamic, elastohydrodynamic, partial, and boundary.

Copyright © 2004 Marcel Dekker, Inc.

3.12 Problems

3.1 Show that for a Gaussian distribution and a zero mean ($z^* = 0$) from the reference line (determined by the M system) the theoretical skewness is zero and the kurtosis is 3.

3.2 Show that the skewness for a deeply pitted surface is less than zero ($\bar{\alpha} < 0$) and that the kurtosis for a relatively flat height distribution approaches zero ($\bar{\beta} \to 0$).

3.3 Find R_a/R_q for a Gaussian distribution with zero mean ($z^* = 0$).

3.4 Prove that $R_a < R_q$.

3.5 What profile would produce an $R_a = R_q$?

3.6 Prove that the kurtosis is greater than or equal to 1.

3.7 What is the skewness-kurtosis inequality? That is, when plotting skewness versus kurtosis, describe the critical curve that separates where the results are allowed and where they are forbidden.

3.8 Show that for a Gaussian distribution with a nonzero mean ($z^* \neq 0$) from the reference line (determined by the M system), the theoretical skewness is 0 and the kurtosis is 3.

3.9 A surface with a triangular sawtooth roughness pattern has a peak-to-valley height of 5 μm. Find the R_a value and the R_q value.

3.10 A precision ball bearing has a race root mean square surface roughness of 0.07 μm and a ball root mean square surface roughness of 0.02 μm. Changing the roughness of the components may give a higher Λ value. Determine which component to smooth if it costs equally as much to halve the roughness of the race as it costs to halve the roughness of the balls. Note that $h_{\min}$=0.2 μm.

3.11 Consider a two-dimensional surface in the shape of a sine wave with an amplitude of 1 μm. Obtain the first four moments of the distribution curve for this surface. Repeat the problem assuming that this surface has encountered wear, so that the top half of the sine wave has been flattened.

3.12 Plot the cumulative distribution function and the Abbott-Firestone curve for the two surfaces in Problem 3.11.

References

Abbot, E. J., and Firestone, F. A. (1933): Specifying Surface Quality. *J. Mech. Eng.*, vol. 55, no. 9, pp. 569-572.

Copyright © 2004 Marcel Dekker, Inc.

ASME (1995a): Surface Texture (Surface Roughness, Waviness, and Lay). American National Standard *ASME B46.1-1995.* New York.

ASME (1995b): Surface Texture Symbols. American National Standard *ASME Y14.36-1995.* New York.

Bhushan, B., ed., (1998): *Handbook of Nano/Micro Tribology.* CRC Press, Boca Raton (FL).

Bhushan B., (2002): *Introduction to Tribology.* Wiley, New York.

Halling, J. (ed.) (1975): *Principles of Tribology.* Macmillan Press, London and Basingstoke.

Halling, J. (1976):*Introduction to Tribology.* Wykeham Publications, London.

Hamrock, B. J., and Dowson, D. (1981): *Ball Bearing Lubrication-The Elastohydrodynamics of Elliptical Contacts.* Wiley-lnterscience, New York.

Persson, U. (1992): "Surface Topography-Speckle Technique and Image Analysis Applied to Surface Roughness Measurement on Machined Surfaces." Ph.D. Thesis. Instrumentation Laboratory, Royal Institute of Technology, Stockholm, TRITA-ILA 92.06.

Sherrington, I., and Smith, E. H. (1988): Modern Measurement Techniques in Surface Metrology: Part I: Stylus Instruments, Electron Microscopy and Non-Optical Comparators. *Wear*, vol. 125, pp. 271-288.

Stout, K.J., Sullican, P.J., Dong, W.P., Mainsah, E., Luo, N., Mathia, T., and Zahouni, H. (1993): The Development of Methods for the Characterization of Roughness in Three Dimensions. Commission of the European Communities, Report EUR 15178 EN.

Tallian, T. E., et al. (1967): On Computing Failure Modes in Rolling Contacts. *ASLE Trans.*, vol. 10. no. 4, pp. 418-435.

Thomas, R.R., (1982): *Rough Surfaces.* Longman, New York.

Copyright © 2004 Marcel Dekker, Inc.